PATENT ABSTRACTS OF JAPAN

(11)Publication number : **05-265496**
(43)Date of publication of application : **15.10.1993**

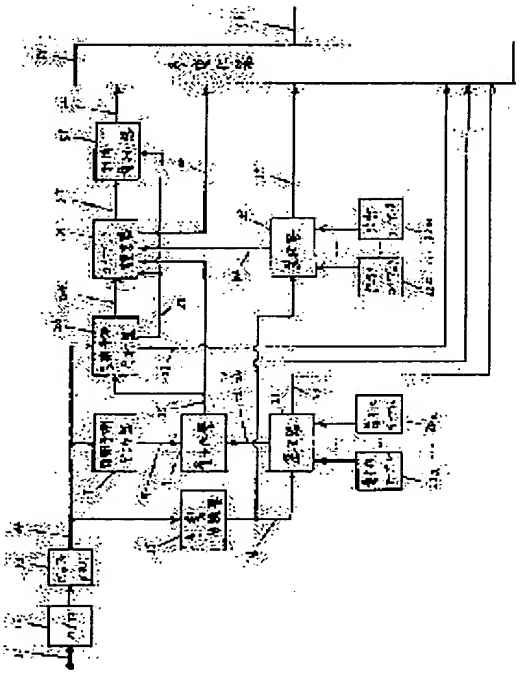
(51)Int.CI. G10L 9/14
G10L 9/18

(21)Application number : **04-061778**
(22)Date of filing : **18.03.1992**

(71)Applicant : **HITACHI LTD**
(72)Inventor : **ASAKAWA YOSHIAKI**
SEKINE HIDETOSHI
SHINADA YASUKO

(54) SPEECH ENCODING METHOD WITH PLURAL CODE BOOKS

(57)Abstract:
PURPOSE: To provide the speech encoding method which can obtain a synthesized speech of high quality even at a low bit rate of ≤4kbps and is relatively small in throughput.
CONSTITUTION: A CELP encoder is equipped with an acoustic classifying unit 15, plural statistical code books 32a...32m, and a statistical code book selector 33, and those statistical code books are switched by the statistical code book selector 33 according to the classification result of the acoustic classifying unit 15 and retrieved by a code book retrieving unit 31. Therefore, the statistical code books 32a-32m are generated by using learning data classified previously by the acoustic classifying unit, so a variation of input speeches can be covered and the quality is improved. Further, only the selected code book is used as a code book retrieval at the time of encoding, so the throughput is reducible.



* NOTICES *

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] this invention relates to the voice coding method especially applicable to the bit rate of 4 or less kbpses by the comparatively few throughput about the suitable voice coding method to obtain a quality synthesized speech by the low bit rate.
[0002]
[Description of the Prior Art] The error with weight of a synthesized speech and fundamental tone voice is evaluated, the voice coding method which determines that a coding parameter will minimize the error and which took in the "analysis-by-synthesis" technique is proposed recently, and it has succeeded in obtaining comparatively good voice quality also in a low bit rate. It is a sign drive-wire type predicting-coding (CELP) method, for example, M.R.Schroeder, as a typical thing. and B.S.Atal:"Code-excited linear prediction(CELP)", Proc.ICASSP 85 (1985. 3) Those with ** and voice quality practical at 4.8kbps are attained. Moreover, many improvement methods of a CELP method are also proposed, for example, the vector-sum drive-wire type predicting-coding (VSELP) method (for example, I.A.Gersonand M.A.Jasiuk:"Vector sum excited linear prediction (VSELP) speech coding at 8kbps", Proc.ICASSP 90 (1990. 4)) is excellent in respect of a throughput, memory space, and bit error resistance.
[0003] On the other hand, digitization of radio gets into stride, and development of the voice coding method of a low bit rate (4 or less kbpses) is desired more from a viewpoint of effective use of frequency. When it is going to form CELP and VSELP into a low bit rate simply, quality degradation becomes large and there is a limitation. Then, the method which switches a drive sound source according to the acoustical property of input voice is proposed.
[0004] the are using multi-pulse to voiced sound and using CELP as such

method, "MPC-CELP" method (Ozawa --) in non-vocal sound Kumagaya : The 3.2 kb/s voice coding method using a multi-pulse and CELP, An electronic-intelligence communication society spring national conference (1990. 3) and the single pulse which controlled a phase and an amplitude by voiced sound, In non-vocal sound, CELP The "SPE-CELP" method to be used (W. Granzow and B.S.Atal:"High-qualitydigital speech at 4 kb/s", Proc.GLOBECOM 90 (1990. 12)), Voice is classified acoustically. For the classification of every, the code book and updating period of CELP The "PS-VXC" method to switch (S.) [Wang] and A.Gersho:"Phonetically-based vectorexcitation coding of speech at 3.6kbps", Proc.ICASSP There is 89 (1989. 5) etc. These coding methods will be called "sound classification coding method" for convenience as contrasted with the uniprocessed type coding method of the conventional CELP etc.

[0005]

[Problem(s) to be Solved by the Invention] Although the above-mentioned sound classification coding method can attain low bit rate-ization by comparatively few quality degradation, it has the following problems.

[0006] Since it is used switching an essentially different coding method (for example, a multi-pulse and CELP), tone quality -- a tone changes -- tends to become unnatural. Even if it forms the 1st purpose of this invention into a low bit rate, it has little degradation of voice quality, and it is offering the coding method with which change of a tone is not noticeable.

[0007] Moreover, the conventional sound classification type coding method needs to possess the coding method with which plurality differs, processing is complicated and the scale at the time of hardware-izing becomes large. The 2nd purpose of this invention is realizing the 1st purpose by the low throughput comparatively.

[0008] Moreover, in order that the conventional sound classification type coding method may determine a coding method uniquely from a sound classification result, the error of a sound classification is connected with degradation of voice quality. The 3rd purpose of this invention is offering the coding method which degradation of voice quality hardly produces, even when a sound classification is mistaken.

[0009]

[Means for Solving the Problem] In order to attain the above-mentioned purpose, in this invention, it has the following meanses. (1) The quantization table reflecting the acoustical feature of input voice on which two or more code books from which a property differs are provided, and quantization properties differ is provided, and it has a means to perform the sound classification of input voice, and a means to choose a quantization table and a code code book based on the result. (2) A VSELP structuring

code book is provided and it has an efficient code book reference means. (3) The code book constituted so that it might not become decisive tone-quality degradation, even if a sound classification was mistaken is provided, and it has the preliminary selection means of a code book. (4) It has a selection means by which a quantization table and a code book choose two or more candidates, and choose the optimal thing out of these all combination.

[0010]

[Function] Since there is various deformation, the operation in typical composition is stated to this invention here.

[0011] First, the operation in the most fundamental composition (the 1st composition) is explained. The voice inputted into the encoder is first divided into a frame and a subframe. In the sound classification section, a classification is made per a frame or subframe according to the acoustical feature of input voice. In the short-term prediction analyzor, a spectrum parameter (short-term prediction coefficient) is extracted for every frame, and a suitable quantization table is chosen and quantized according to the result of the aforementioned sound classification. Next, in the long-term prediction analyzor, the optimal long-term prediction lug and optimal gain are searched for from an adaptation code book per subframe. In the code book reference section, out of two or more statistics code books, one or more code books are chosen according to the result of the aforementioned sound classification, and the optimal code vector and optimal gain are searched for by searching this. In addition, in long-term prediction analysis or code book reference, as the error of weighting ****, fundamental tone voice, and a synthesized speech is minimized, a lug and a code vector are chosen by the aforementioned short-term prediction coefficient.

[0012] It is quantized, and encodes with the index of a long-term prediction lug, the selected statistics code book number, and a code vector, and the spectrum parameter and gain which were searched for as mentioned above are transmitted to a decoder as a transmission parameter.

[0013] In a decoder, a drive sound source is calculated from the above-mentioned transmission parameter, and decryption voice is obtained by being inputted into the synthetic filter which makes a short-term prediction coefficient a filter factor.

[0014] Moreover, in another composition of this invention, without performing a sound classification, two or more statistics code books of all are searched, and the optimal code vector is chosen. The other operation is the same as that of the case of the composition of the above 1st.

[0015] Moreover, in still more nearly another composition of this invention, the statistics code book is divided into two or more subsets, and

limits the subset of the code book to search according to a sound classification result. The other operation is the same as that of the case of the composition of the above 1st.

[0016]

[Example] Hereafter, the example of this invention is explained using a drawing. The block diagram of the coding section of the 1st example of this invention is shown in drawing 1, and the block diagram of the decryption section is shown in drawing 2 .

[0017] Since this invention is based on the sign drive type voice coding (CELP) method, it explains the principle of a CELP method first in advance of explanation of a concrete example. Drawing 3 is the principle view of the drive sound-source determination in the coding section of CELP. Let the weighted sum which multiplied by it and added gain to each be a drive sound source in this drawing using the long-term prediction vector 107 which is the output of the adaptation code book 104 as a component showing the periodicity of a sound source, and the code vector which is the output of a statistics code book as components other than periodicity. In addition, let the random nature and noise nature of a sound source be a code vector as components other than the periodicity of a sound source in the illustrated example, respectively. Therefore, each gain 111 and 112 is multiplied by it and added to the code vectors 108 and 109 which are the outputs of two statistics code books 105 and 106 corresponding to each as a statistics code book. The kind of statistics code book also has one case.

[0018] Reference of the code book for acquiring the optimal drive sound source is made as follows. Although a drive sound source whose synthesized speech which generally inputs a drive sound source into a synthetic filter, and is obtained corresponds with fundamental tone voice (input voice) should just be acquired, it is accompanied by a certain error (quantumization noise) in fact. Therefore, although what is necessary will be just to determine that a drive sound source will minimize this error, it is known that human being's acoustic-sense property cannot necessarily take correspondence of the subjectivity quality of the amount of errors and voice. Then, it is common to use the error which carried out weighting so that correspondence with an acoustic-sense property might become good. Acoustic-sense weighting is indicated by the following reference, for example. B.S. Atal and J.R.Remde:"A new model ofLPC excitation for producing natural-sounding speech at low bitrates", Proc.ICASSP 82 (1982. 5).

[0019] In order to evaluate this acoustic-sense weighting error, the drive sound source 113 is inputted into the weighting composition filter 115, and obtains the weighting synthesized speech 115. The weighting input voice

118 is obtained through the weighting filter 117, and the input voice 116 also takes a difference with the weighting synthesized speech 115, and acquires the weighting error wave 119. A weighting error wave has the square sum calculated over the error evaluation section in the square error calculation section 120, and the weighting square error 121 is acquired. As mentioned above, since a drive sound source is the load sum of a long-term prediction vector and a statistics code vector, the determination of a drive sound source results in the determination of a code-vector index which decides which code vector to choose from each code book. Namely, what is necessary is to compute the weighting square error 121 by changing the long-term prediction lug 101 and the code-vector indexes 102 and 103 one by one, and just to choose that from which a weighting error serves as the minimum in the error minimization section 122. Such a drive sound-source determining method is called "analysis-by-synthesis" method. If it is similarly going to optimize the index of a long-term prediction lug and a statistics code vector, performing an above-mentioned procedure faithfully, namely, evaluating a weighting error each time, since it will become a huge throughput, technique, such as optimization, is used serially in fact.

[0020] Drawing 1 is the block diagram of the voice coding section of the 1st example of this invention. The coding section divides roughly and consists of the voice input section, the short-term prediction analyzor, the long-term prediction analyzor, the code book reference section, and the gain quantization section. Drawing 2 is the block diagram of the decryption section of the 1st example of this invention. The decryption section divides roughly and consists of the parameter decryption section, the drive sound-source generation section, and the synthesized-speech output section. Hereafter, the outline of operation of the 1st example is explained.

[0021] The analog input voice 11 is changed into a digital signal by A-D converter 12, and is once stored in buffer memory 13. The sampling frequency of an AD translation is usually 8kHz. The sound sorter 15 reads the digital voice data 14 of frame length or subframe length from buffer memory 13, and classifies it based on the acoustical feature of input voice. Frame length and subframe length are 40ms (320 samples) and 8ms (64 samples) grade, respectively. Moreover, a sound classification is classified into vocality, friction nature, etc. like the after-mentioned. A sound classification result is outputted as a sound classification flag 16.

[0022] The short-term prediction analyzer 17 reads the voice data 14 of analysis frame length from buffer memory 13, and outputs the short-term prediction parameter 18.

[0023] The short-term prediction parameter 18 is quantized in the short-term prediction parameter quantizer 19. Actual quantization is presented

with the quantization table 22 chosen by the short-term prediction parameter quantization table selector 21 with reference to the sound classification flag 16 from 20n from two or more [here] quantization table 20a. The number of the selected quantization table is outputted as a short-term prediction parameter quantization index 24, and the quantization sign as a quantization table index 23 is outputted as a transmission parameter, respectively. These two indexes cannot be overemphasized by being collected into one. Moreover, the quantization value 25 of a short-term prediction parameter is sent out after the next step.

[0024] In the long-term prediction analyzer 26, the long-term prediction lug showing audio periodicity which is a parameter is extracted, and the index 27 and the long-term prediction vector 28 of a long-term prediction lug are outputted.

[0025] The noise component of a sound source is searched with the statistics code book reference machine 31 from a code book. The statistics code book selector 33 out of 32m chooses the code book 34 for reference from two or more code book 32a with reference to the sound classification flag 16. The index 35 of a code book and the index 36 of a code vector are outputted as a transmission parameter. Moreover, the code vector 37 obtained as a reference result is sent out to the gain quantizer 51.

[0026] In the gain quantizer 51, the gain of the long-term prediction vector 28 and a code vector 37 is quantized, and the gain index 52 is outputted.

[0027] With the multiplexing vessel 54, the various indexes 23, 24, 27, 35, 36, and 52 which are transmission parameters are multiplexed, and it is sent out to a transmission line 55.

[0028] Next, the outline of operation of the decryption section is explained using drawing 2 .

[0029] The transmission parameter received from the transmission line 55 is separated into the various parameters 61, 62, 63, 64, 65, and 66 by the demultiplexing machine 56. If these parameters do not receive a bit error on a transmission line, they have the same value as the parameters 23, 24, 27, 35, and 36 sent out in the coding section.

[0030] In the short-term prediction parameter decoder 72, with reference to the quantization table index 61, a quantization table is chosen by the short-term prediction parameter quantization table selector 71 from 70n from two or more short-term prediction parameter quantization table 70a, and the decode of the short-term prediction parameter (quantization value) 73 is carried out from this selected quantization table based on the short-term prediction parameter quantization index 62. In addition, 70n cannot be overemphasized by that it is equal to 20n from quantization table 20a in the coding section from quantization table 70a.

[0031] In the long-term prediction lug decoder 74, the decode of the long-term prediction lug 75 is carried out based on the long-term prediction lug index 63.
[0032] In the quantization gain decoder 76, gain 77 is decoded based on the gain index 66.
[0033] In the statistics code-vector decoder 80, with reference to the code book index 64, a code book is chosen by the statistics code book selector 79 from 78m from two or more statistics code book 78a, and the decode of the code vector 81 is carried out based on the code-vector index 65 from this selected code book. 78m is equal to 32m from statistics code book 32a in the coding section from statistics code book 78a like a quantization table.
[0034] With the drive sound-source generation vessel 92, based on the long-term prediction lug 75, the long-term prediction vector 91 is first read from the adaptation code book 90, and the drive sound source 93 is generated for every subframe with a code vector 81 and gain 77. The drive sound source 93 is used also for renewal of the adaptation code book 90 while it is inputted into the synthetic filter 94. Although the adaptation code book is not specified by the block diagram of the coding section of drawing 1 , as explanation of the below-mentioned long-term prediction analyzer 26 shows, the same thing possesses it also in the coding section.
[0035] The synthetic filter 94 is a linear-prediction filter which makes a filter factor the parameter drawn from the quantization short-term prediction parameter 73 or it, and carries out the synthetic output of the digital synthesized speech 95.
[0036] The digital synthesized speech 95 is continuously sent out to DA converter 97 through buffer memory 96, and is changed into the analog synthesized speech 98.
[0037] In the above, since the outline was explained, the detailed function of the main portions of the 1st example is explained below.
[0038] The sound sorter 15 calculates a physical parameter from the voice data 14 of frame length or subframe length, and classifies the voice of the section into two or more categories according to the logic judging of those parameter value. The sound classification method itself is well-known technology, for example, an example is indicated by 4.8 kb/s multi-pulse voice coding method" using an Ozawa:" various sound source, and the Acoustical Society of Japan lecture collected works (1989. 3). The block diagram at the time of constituting as a sound sorter is shown in drawing 4 . a physical parameter -- the energy calculation section 151, the energy rate-of-change calculation section 153, the maximum correlation calculation section 155, the prediction gain calculation section 157, and a logarithm -- it is calculated by the arca-ratio calculation section 159, and each physical

quantity 152, 154, 156, 158, and 160 is inputted into the logic judging section 161 Refer to the aforementioned reference for the concrete calculation method of each physical quantity. The judgment logic of the logic judging section 161 is as being shown in drawing 5 , and is classified into four kinds, vocality, nasal nature, rupture and transient nature, and friction nature, after this. Moreover, the artificers of this invention have proposed the another sound classification method. The block diagram is shown in drawing 6 . Physical parameters are energy 152, the energy rate of change 154, and the maximum correlation value 156. The judgment logic of the logic judging section 162 is shown in drawing 7 . It has classified into four kinds, a vowel and nasal nature, a standup, falling, and others, according to this example. Although a sound classification is performed per a frame unit or subframe, energy rate of change should just compute change of the difference of the frame energy of a front frame, and the frame energy of the present frame, or the energy for every subframe, for example, when computing per frame. Moreover, what is necessary is to divide the energy difference of the adjoining subframe, or a subframe into half order further, and just to detect the difference of the energy of each, when computing per subframe.

[0039] The short-term prediction analyzer 17 extracts the short-term prediction parameter 18 showing an audio spectral envelope from voice data 14 for every frame. Most generally, the short-term prediction parameter 18 is linear predictor coefficients or an equivalent parameter drawn from it, and specifically has a partial auto correlation coefficient (a PARCOR coefficient, reflection coefficient), a line spectrum pair (loop-splice-plate parameter), etc. as the derivation method of linear predictor coefficients -- the repeating method (Saito --) of Durbin-Levinson General it will be introduced in the Nakada work, "the foundation of speech information processing", Ohm-Sha Ltd., and Showa 56 -- **** -- the derivation method of a reflection coefficient It is a FLAT algorithm (Research & Development Center for Radio System decision) besides the above. it is indicated by "digital method car telephone system standard RCR STD-27" (it abbreviates to "RCR specification document" hereafter) -- **** -- the LeRoux method (above Saito, the Nakada work, and writing publication) etc. is proposed moreover, the conversion method from linear predictor coefficients to a loop-splice-plate parameter -- the above Saito and written by Nakada -- it is indicated by writing

[0040] It is quantized by the linear-prediction parameter quantizer 19, and the linear-prediction parameter 18 is changed into the quantization value 25. Although the formation of a scalar quantity child and a vector quantization are used by the number of bits by which quantization is

permitted, the feature of this invention is providing two or more quantization tables 20a, ..., 20n. Since the distribution of linear-prediction parameter value changes with audio features, the efficient quantization of it is attained by using the quantization table according to the audio feature. Then, in the short-term prediction parameter quantization table selector 21, the quantization table 22 for reference is chosen with reference to the sound classification flag 16. The quantization table selector 21 matches a sound classification result and the quantization table for reference, and serves as table reference form. Usually, although one quantization table is chosen to one sound classification, depending on the number of quantization tables, or the category of a sound classification, two or more quantization tables may be chosen as a quantization table 22 for reference. It actually quantizes, and if the quantization value 25 from which a quantumization noise serves as the minimum is decided, the index 23 of the quantization table then used and the sign (quantization index) 24 of a quantization level will be outputted as a transmission parameter. in addition, although two transmission parameters 23 and 24 were separate-boiled for convenience and being indicated, summarizing both and being made to one parameter cannot be overemphasized

[0041] Although [the above explanation] the quantization table 20 has more than one, in this invention, a quantization table can be provided only one. In this case, a quantizer 19 searches the direct quantization table 20 through a selector 21.

[0042] Next, the long-term prediction analyzer 26 is explained. It is considered that long-term prediction analysis is reference of an adaptation code book, and a long-term prediction lug (index of an adaptation code book) is chosen by minimization of the acoustic-sense weighting error of a synthetic wave and fundamental tone voice as explained first. Here, the case where it searches sequentially is explained to be a statistics code book. That is, the output of a statistics code book is assumed to be 0, and determines the optimal long-term prediction vector 28.

[0043] In order to compute a weighting error, in the acoustic-sense weighting section 261, weighting is first made by the input voice 14, and the weighting voice 262 is obtained. Although a weighting filter consists of quantization short-term prediction parameters 25, the concrete form is as follows.

[0044]

[Equation 1]

$$W(z) = \frac{1 - \sum_{i=1}^{Np} \alpha_i z^{-i}}{1 - \sum_{i=1}^{Np} \alpha_i \lambda^i z^{-i}} \qquad (数1)$$

[0045] alphai is [Np=10 and lambda of a filter factor (linear predictor coefficients) and Np] usually lambda= 0.8 with a weighting parameter in the number of filters here.
[0046] Generally, although the output of a synthetic filter is influenced of the past state, in order to cut down the amount of operations here, it removes the influence of the past synthetic filter from the weighting voice 262 beforehand. That is, in zero input response calculation / subtraction section 263, the zero input response of a weighting composition filter is calculated, it subtracts from the weighting voice 262, and the weighting voice 264 which removed the influence of past is obtained. The synthetic filter used here is as follows.
[0047]
[Equation 2]

$$H(z) = \frac{1}{1 - \sum_{i=1}^{Np} \alpha_i \lambda^i z^{-i}} \qquad (数2)$$

[0048] This synthetic filter is the point that the point containing the weighting parameter lambda differs from the synthetic filter by the side of decode.
[0049] weighting composition of the code vector 268 read from the adaptation code book 267 on the other hand corresponding to the long-term prediction lug set as the object of reference -- the long-term prediction calculus-of-vectors section 269 -- setting -- the impulse response of a weighting composition filter -- collapsing -- it realizes Thus, since it does not depend on the state of the past of a synthetic filter for the obtained synthetic output (long-term prediction vector) 270, it is called a zero state response. The impulse response of a weighting composition filter is beforehand calculated in the impulse response calculation section 265 here, using the quantization value 25 of a short-term prediction parameter as alpha of (several 2). The long-term prediction vector 270 over each lug in the reference range is calculated, correlation with the weighting voice 264 is calculated in the lug selection section 271, and the long-term prediction lug index 27 which quantized the long-term prediction vector 28 which gives mutually related maximum (it is the optimal), and the long-term prediction lug at that time is outputted. Please refer to the above-mentioned

RCR specification document about the detail of long-term prediction analytical method, or the technique for the amount curtailment of operations.

[0050] Next, the statistics code book reference machine 31 is explained. The feature of this invention possesses 32m from two or more kinds code book 32a, and is that it changes and uses them based on the result of a sound classification. As for the reason carried out in this way, it is raised for every audio feature that the features of the remainder wave (it is a signal equivalent to the sound-source wave inputted into a synthetic filter, and obtained by audio analysis) also differ. The code vector of a statistics code book serves as an almost random noise component, when it is thought that the remainder wave after predicting voice short-term prediction and over a long period of time is approximated and short-term prediction and long-term prediction are made ideally. However, it is the influence of quantization etc. actually, and short-term prediction and long-term prediction are not performed ideally, but, as a result, the audio feature comes to remain also in the remainder. Therefore, it is effective to make the feature reflect for every audio category, and to create a statistics code book because of quality improvement. Moreover, it is effective to limit beforehand the code book which should be searched also in respect of code book reference throughput curtailment.

[0051] The block diagram of a code book reference machine is shown in drawing 9 . The statistics code book selector 33 chooses the code book 34 for reference from two or more statistics code books 32a, ..., 32m with reference to the sound classification flag 16. In the zero state response calculation section 311, as for the code vector read from the code book 34 for reference one by one, the zero state response 312 is calculated like long-term prediction analysis using the quantization value 25 of a short-term prediction parameter. The zero state response 312 is orthogonalized with the long-term prediction vector 28 in the orthogonalization section 313. Gram Schmidt's method of orthogonalizing etc. is used for orthogonalization. In code book reference, although orthogonalization is not indispensable processing, it is effective in reducing the performance degradation by the sequential retrieval with a long-term prediction vector. The orthogonalized zero state response 314 is a square error [0052] with the weighting voice 264 which removed the influence of the past of the synthetic filter which was inputted into the reference section 315 and called for by long-term prediction analysis.

[Equation 3]

$$E = \sum_{n=0}^{N-1} \{ p(n) - \gamma f_i(n) \}^2 \qquad (数3)$$

[0053] It ******. Gain and N of a code vector [finishing / here / the above-mentioned weighting voice and f (n) are orthogonalized for p (n) here, and / filtering] and gamma are the measurement sizes in a subframe. Moreover, i is the index of a code vector. The index 36 of the code vector which gives the minimum value of a square error is outputted, and the zero state response corresponding to an index 36 is outputted as a code vector (the optimal) 37 in the code-vector calculation section 316.
[0054] Above, the statistics code book is explained like the usual CELP encoder as a set of the code vector which constitutes a drive sound source. It is possible to adopt a set of a VSELP type basis vector as code book structure in this invention. In this case, it can consider that the binary load sum of all basis vectors is a code vector, and reference of a code book results in how to combine a binary load. The code vector of the usual CELP type encoder and the basis vector of a VSELP encoder are matched by the following formula.
[0055]
[Equation 4]

$$U_i(n) = \sum_{m=1}^{M} \theta_{im} V_m(n) \qquad (数4)$$

[0056] Vm (n) shows the m-th basis vector here, and M expresses the number of basis vectors, 9 [for example,], to it. n is a sample number in a subframe, and if subframe length considers for example, as 64 samples, n will take the value of 0 to 63. Ui (n) is the code vector of the m-th power individual (in the case of M= 9, they are 512 pieces) of 2 generated by the linear combination of M basis vectors, and i takes the value to the m-th power -1 of 0 to 2. Weight thetaim of linear combination will take thetaim=+1 and the binary value called thetaim=-1 if it becomes zero, if m bits of a symbolic language i become one. That is, a code vector Ui (n) is generated by the total combination of addition of M basis vectors, or subtraction.
[0057] About code book reference of a VSELP encoder, since it is stated to the above-mentioned RCR specification document in detail, it omits here. Drawing 10 shows the block diagram at the time of taking in VSELP type code book structure to this invention. The code book 34 which serves as a candidate for reference by the selector 33 is chosen from two or more basis vector code book 32'a, ..., 32'm. Although zero state response calculation and orthogonalization are the same as that of explanation by drawing 9 , it differs in that objects are not a code vector but M basis vectors. The zero state response 314 of each basis vector orthogonalized by the long-term

prediction vector 28 is inputted into the load sum calculation section 317, load thetaim stored in the load storing section 318 is read one by one, and a code vector is calculated according to (several 4). This is equivalent to what generated the code vector from the basis vector beforehand, zero-state-answered and orthogonalized this. About the code vector 320 corresponding to the load as the m-th power of 2, a square error with the input voice 264 which had the eclipse zero input response with weight subtracted is evaluated by the code-vector reference section 315. The sign 36 of weight which gives the minimum value of a square error is outputted as a transmission parameter, and is simultaneously inputted also into the code-vector calculation section 316. In the code-vector calculation section 316, the code vector 37 which is not orthogonalized by the long-term prediction vector is calculated by reading the load 319 corresponding to the sign 36 which gives the minimum value of a square error from the load storing section 318, and being based to the zero state response 312 of a basis vector (several 4).

[0058] The creation method of a code book is explained in relation to code book reference. As a method of designing a code book (study), what is depended on a closed-loop method is proposed recently. A closed-loop method performs the same processing as actual coding using the study data of a code vector, and it corrects study data so that an error may decrease. the code book learning method of a CELP encoder -- for example, -- G.Davidson and M.Yong and A.Gersho:" -- Real-time vector excitation coding of speech at 4800 bps"Proc.ICASSP It is stated to 87 (1987). moreover, the learning method of the basis vector of VSELP -- I.A.Gerson: -- "Vector sum excited linear prediction(VSELP) speech coding for Japan digital cellular", and a **** technical report, RCS 90-20 (1990. 11) and plastic watt and Itakura: -- it is told to "the design of the drive sound-source wave code book in linear predictive coding", and a **** technical report and SP 90-53 (1990) In this invention, although a Prior art is used for the learning method of a code book or a basis vector, only the number of categories according to the sound classification has the feature in carrying out code book possession. Drawing 11 is a block diagram for creating the study data for code book study corresponding to the sound classification. Study data are made from sufficient quantity of voice data so that it can respond also to an unspecified speaker and the content of phonation enough. Since code-vector (or it is the same as that of basis vector and the following) length is in agreement with subframe length, he reads the digital voice data 14 for study by subframe length, and inputs this into the sound sorter 15 and the parameter calculation section 291. In the parameter calculation section, on data required for study of a code book, and a

concrete target, a linear-prediction parameter, an impulse response, a long-term prediction vector, etc. are calculated, these parameters 292 are changed, and it outputs to the section 293. On the other hand, the sound sorter 15 has the same function as being used by the voice coder, extracts the acoustical feature of voice data 14, changes a classification and the sound classification flag 16 to the category of a predetermined number, and outputs them to the section 293. In the change section 293, the parameter 292 calculated in the parameter calculation section 291 according to the sound classification flag 16 is distributed to the study data storage sections 294a, ..., 294m.

[0059] Next, it returns to drawing 1 and the gain quantizer 51 is explained. This load is gain although a drive sound source is the load sum of the long-term prediction vector C0 and the statistics code vector C1 like the following formula (there is not each of C0 and C1 eclipse **** with weight).

[0060]

[Equation 5]

$$ex(n) = \beta c_0(n) + \gamma c_1(n) \qquad (数5)$$

[0061] Gain can be searched for by carrying out the partial differential of the error evaluation formula in long-term prediction analysis or code book reference. However, in this example, since the sequential retrieval is performed, after the optimal gain searches for each optimal vector, it is necessary to search for it by the partial differential of the following overall-error evaluation formula.

[0062]

[Equation 6]

$$E = \sum_{n=0}^{N-1} \{p(n) - \beta c_0'(n) - \gamma c_1'(n)\}^2 \qquad (数6)$$

[0063] C0' and C1' are an eclipse ****** prediction vector with weight, and a statistics code vector here. them after doing in this way and searching for optimal gain in a gain quantizer -- the formation of a scalar quantity child -- or it vector-quantizes and the gain quantization index 52 is outputted Moreover, it is also possible to quantize gain by the closed-loop method. This reads the quantization value (candidate) of gain from a quantization table one by one, evaluates an error using this, and is taken as a quantization result with the quantization value which gives the minimum value with error. The example of this method is indicated by the above-mentioned RCR specification document.

[0064] A transmission parameter becomes as follows as a result of coding processing of this example explained above. They are six parameters of the quantization table index 23, the short-term prediction parameter quantization index 24, the long-term prediction lug index 27, the code book index 35, the code-vector index 36, and the gain quantization index 52. These are multiplexed with the multiplexing vessel 54 and sent out to a transmission line 55.
[0065] Next, it returns to drawing 2 and the decryption section of this example is explained.
[0066] If received from a transmission line 55, in the demultiplexing machine 56, demultiplexing of the transmission parameter will be carried out to the quantization table index 61, the short-term prediction parameter quantization index 62, the long-term prediction lug index 63, the code book index 64, the code-vector index 65, and the gain quantization index 66.
[0067] The 1st phase of decryption processing is a decryption of each parameter value. The short-term prediction parameter quantization table selector 71 chooses the quantization table chosen at the time of coding based on the quantization table index 61 from two or more quantization tables 71a, ..., 70n, and sends it out to the short-term prediction parameter decoder 72. In this decoder 72, the decode of the short-term prediction parameter value 73 is carried out based on the short-term prediction parameter quantization index 62. By the long-term prediction lug decoder 74, the long-term prediction lug 75 is similarly decoded based on the long-term prediction lug index 63. In the gain decoder 76, the quantization gain 77 is decoded based on the gain quantization index 66. The statistics code book selector 79 chooses the statistics code book chosen at the time of coding based on the code book index 64 from two or more statistics code books 78a, ..., 78m, and sends it out to the statistics code-vector decoder 80. In this decoder 80, the decode of the code vector 81 is carried out based on the code-vector index 65.
[0068] The 2nd phase of decryption processing is generation of a drive sound source. With the drive sound-source generation vessel 92, as shown in (several 5), gain 77 is multiplied by it and added to the long-term prediction vector 91 read from the adaptation code book 90 corresponding to the long-term prediction lug 75, and a code vector 81, and the drive sound source 93 is generated. The drive sound source 93 is used also for the renewal of a state of the adaptation code book 90 while it is inputted into the synthetic filter 94.
[0069] The stage of the last of decryption processing is speech synthesis. With the synthetic filter 94, the short-term prediction parameter 73 by which decode was carried out by the short-term prediction parameter

decoder 72 is made into a filter factor, and the synthetic output of the digital synthesized speech 95 is carried out by inputting the drive sound source 93. The digital synthesized speech 95 is continuously sent to a DA converter through buffer memory 96, and is changed into the analog synthesized speech 98.

[0070] Above, operation from the voice input of the 1st example of this invention to coding, a decryption, and a voice output was explained. Especially audio frame energy (power) was not mentioned in the above explanation. It is more advantageous to normalize with frame energy beforehand, in order to stop the dynamic range of gain, if quantization of gain is taken into consideration, although this is because frame energy is reflected in the gain of a drive sound source. Since frame energy is easily searched for at the time of calculation of a linear-prediction parameter, it quantizes separately and frame energy transmits the index. The example of the bit allocation at the time of doing in this way is shown below.

[0071] It is set as 8kHz and frame length, and subframe length is set to 8ms (64 samples) for a sampling frequency for 40ms (320 samples). Frame energy and a linear-prediction parameter shall be updated per frame, and other parameters shall be updated per subframe. In addition, it is more effective in upgrading of a synthesized speech to have interpolated frame energy and the linear-prediction parameter per subframe, and to use them. When it has the quantization table of a short-term prediction parameter two kinds, a quantization table index (change flag) is 1 bit. A quantization index becomes 20 bits if quantization performs a 20-bit two-step vector quantization. Frame energy is formed into a scalar quantity child by 5 bits. Therefore, the number of transmitted bits per frame is 26 bits.

[0072] The index of a long-term prediction lug of the parameter of a subframe unit is 7 bits, and, as for this, the range of a long-term prediction lug corresponds to 146 samples (55Hz) from 19 samples (421Hz). If four kinds of statistics code books are provided, 8 bits (256 code vectors), then the code-vector index of a code book index are 8 bits about 2 bits and code book size. Gain vector-quantizes the thing to a long-term prediction vector, and the thing to a statistics code vector, and expresses them with 7 bits. Therefore, the number of transmitted bits per subframe becomes 24 bits. A total bit rate is set to 3650bps by the above. In this case, the sound sorter is outputting the classification flag for every subframe for every frame for statistics code book selection for selection of a short-term prediction parameter quantizer. However, the output interval of 1 time, then a sound classification flag also becomes every two subframes at two subframes, and a bit rate reduces the change of a statistics code book further.

[0073] As explained above, in the 1st example of this invention, by the

throughput almost equivalent to the conventional CELP or VSELP, degradation of voice quality is suppressed and low bit rate-ization is attained.

[0074] Next, the 2nd example of this invention is explained. The coding section is shown in drawing 12 and the decryption section is shown in drawing 13 . As the feature of this example is shown in drawing 12 , reference of a statistics code book is two steps, and the all are a certain things [that are, and crawl and or / gap / possesses two or more code books]. as it comes out so and was also with the conventional CELP encoder of drawing 3 , it is for raising the order of approximation of noise components other than the periodicity of a drive sound source to search a statistics code book in two stages, and it is performing two-step vector quantization -- it can be rich and can also make

[0075] The coding section of this example serves as the form where the 2nd statistics code book reference machine 41 was inserted between the 1st statistics code book reference machine 31 and the gain quantizer 51, in the 1st example of drawing 1 , as shown in drawing 12 . Hereafter, although this example is explained, explanation is omitted about an intersection with the 1st example, and only a 2nd code book reference machine-related portion is explained.

[0076] As shown in drawing 14 , the code book 44 for reference is chosen by the statistics code book selector 43 from two or more code books 42a, ..., 42l. Although it is fundamentally [the structure of the 2nd statistics code book reference machine 41 / as the 1st statistics code book reference machine 31 of drawing 9] the same, it differs in that the zero state response 412 is orthogonalized also not only to the long-term prediction vector 28 but to the 1st code vector 37 in the orthogonalization section 413. The zero state response 414 which the zero state response by which the zero state response 412 was first orthogonalized by the long-term prediction vector 28, and was specifically orthogonalized by this long-term prediction vector was further orthogonalized to the 1st statistics code vector 37, and was orthogonalized by two stages is inputted into the reference section 415. The output of the 2nd statistics code-vector reference machine 41 is the 2nd code-vector index 46 and 2nd code vector 47.

[0077] In the gain quantizer 51, the gain over the long-term prediction vector 28, the 1st code vector 37, and the 2nd code vector 47 is searched for and quantized, and the gain quantization index 52 is outputted.

[0078] The block diagram of the decryption section of the 2nd example of this invention is shown in drawing 13 . It has the 2nd statistics code book 82a, ..., 82l. and the structure where the 2nd statistics code book selector 83 and the 2nd code-vector decoder 84 were added at the decryption section

(drawing 2) of the 1st example. From the gain decoder 76, the gain over three vectors, a long-term prediction vector, the 1st code vector, and the 2nd code vector, is outputted as gain 77. With the drive sound-source generation vessel 92, gain is multiplied by it and added to the long-term prediction vector 91, the 1st code vector 81, and the 2nd code vector 85, respectively, and a drive sound source is generated.

[0079] In the 2nd example, although the throughput of the 2nd statistics code book reference and the bit rate of the 2nd code book index and the 2nd code-vector index increase, a quality synthesized speech can be obtained compared with the 1st example.

[0080] Next, the 3rd example of this invention is explained. The coding section is shown in drawing 15 . The decryption section is the same as the decryption section (drawing 2) of the 1st example. The feature of this example is providing the error evaluation machine 53, as shown in drawing 15 . That is, in the coding section, it has two or more quantization values of a short-term prediction parameter, and code vectors of a statistics code book as a candidate, respectively, a weighting error is calculated about those total combination, and the index of the combination which minimizes the error is considered as the final output of an encoder. This tends to reduce degradation from the joint optimization of optimization serially. An effect increases more by taking out two or more candidates not only about a short-term prediction parameter and a statistics code vector but about a long-term prediction vector. Hereafter, the main portions of this example are explained.

[0081] In the short-term prediction parameter quantization table selector 21, two or more candidates are chosen as a quantization table for reference. This should just assign two or more quantization tables to the value of the sound classification flag 16. In the short-term prediction parameter quantizer 19, the short-term prediction parameter value 25 quantized using each quantization table and the quantization index 24 at that time are outputted. When the quantization table is assigned only for one to the value of the sound classification flag 16, the quantumization noise when quantizing using the quantization table outputs the candidate of the predetermined number to small order.

[0082] in the long-term prediction analyzer 26, the quantization value 25 of two or more short-term prediction parameters is alike, respectively, and it receives, and asks for a long-term prediction lug, and the long-term prediction vector 28 is outputted For example, if the quantization value 25 of a short-term prediction parameter has two candidates, two long-term prediction vectors are also acquired. the quantization value of a short-term prediction parameter also boils a long-term prediction lug, respectively, it

receives, and if a candidate [two or more (for example, two pieces)] is taken out, as a combination, four kinds will be made at this time Hereafter, also in reference of a statistics code book, it is the same, and the candidate of further two or more statistics code vectors is taken out to the combination of the candidate before two or more them. Eight kinds of combination can do the number of candidates in 2, then all. In a gain quantizer, the respectively optimal gain is searched for from eight kinds of this combination, and the index is outputted.

[0083] With the error evaluation vessel 53, about eight kinds of such combination, each weighting square error is computed, combination which gives the minimum value is made into a final coding result, and the following parameters are outputted. quantization -- a table -- an index -- 23 -- ' -- a short period -- prediction -- a parameter -- quantization -- an index -- 24 -- ' -- a long period of time -- prediction -- a lug -- an index -- 27 -- ' -- a code book -- an index -- 35 -- ' -- a code vector -- an index -- 36 -- ' -- gain -- quantization -- an index -- 52 -- ' -- it is .

[0084] In the decryption section, each parameter value is decoded from these transmission parameters, and, finally a synthesized speech is obtained.

[0085] although a throughput and middle data storage capacity increase in this example compared with the 1st example in order to take out two or more candidate outputs in each processing section and to carry out error evaluation to the combination, the quality of a synthesized speech is boiled markedly and improves

[0086] Next, the 4th example of this invention is explained. The coding section is shown in drawing 16 . The decryption section is the same as the decryption section (drawing 2) of the 1st example. Although the quantization table of a short-term prediction parameter and the point of providing two or more statistics code books, respectively are the same as the 1st to 3rd example at this example, the feature is in the point of performing these selections, without being based on the classification result of a sound sorter. That is, in quantization of a short-term prediction parameter, it quantizes using two or more quantization tables of all, and that from which a quantization error serves as the minimum is chosen. Moreover, two or more statistics code books of all are searched with reference of a code book, and that from which a weighting error becomes the minimum is chosen by it. Although this creates the quantization table or the statistics code book so that an audio variation may be covered based on a sound classification, what an error minimization norm determines is meant at the time of quantization or reference.

[0087] Since quantization of a short-term prediction parameter and

reference of a statistics code book turn into all search, although a throughput increases compared with the 1st example according to this example, a synthesized speech with the good object scale (for example, SEGUMENTARU SN ratio) showing voice quality is obtained.
[0088] Next, the 5th example of this invention is explained. The coding section is shown in drawing 17 . In this example, a sound classification is not performed like the 4th example. Except it, it is the same as the 2nd example, and the statistics code book is searched in two stages. The decryption section is the same as the decryption section (drawing 13) of the 2nd example.
[0089] The effect of this example is the point that a synthesized speech with the object scale (for example, SEGUMENTARU SN ratio) showing voice quality good although a throughput increases compared with the 2nd example is obtained like the case of the 4th example.
[0090] Next, the 6th example of this invention is explained. The coding section is shown in drawing 18 . In this example, a sound classification is not performed like the 4th example. Except it, it has the error evaluation machine 53 like the 3rd example, and that from which a weighting error serves as the minimum among the combination of each candidates of two or more of the quantization value 25 of a short-term prediction parameter, the long-term prediction vector 28, and the code vector 37 of a statistics code book is determined. The decryption section is the same as the decryption section (drawing 2) of the 1st example.
[0091] In the 3rd example, although it was also more possible than the number of quantization tables actually provided by the short-term prediction parameter quantization table selector 21 or the statistics code book selector 33, and the number of statistics code books to have narrowed down the number of candidates, the candidate of only the actually provided number of quantization tables or the number of statistics code books will go up by this example. It is possible to extract a final candidate from the inside on the basis of a quantumization noise or a weighting error, of course.
[0092] like the 3rd example, although a throughput and middle data storage capacity increase compared with the 1st example, the quality of a synthesized speech has the effect of this example in the point which is markedly alike and improves
[0093] As mentioned above, in the 6th example, two or more statistics code books use as it is what was created corresponding to the sound classification from the 1st example. However, these code books are not completely independent and are not necessarily separated mutually. That is, if there is what is similar to a component (code vector), or duplication among two or more code books, a compacter code book can be constituted

by unifying two or more code books and clustering again. In this case, it can be considered that the code book before integration is the subset of the code book after integration. Therefore, instead of two or more statistics code books provided in the old example, the code book after integration is used, and a statistics code book selector can specify the subset of an integrated code book, and let it be a code book for reference.

[0094] The relation between an integrated code book and the code book for reference is shown in drawing 18 . The integrated code book 321 re-clusters the individual code book created corresponding to the sound classification, and unifies it. The code book 34 for reference is the subset of the integrated code book 321. Fundamentally, the function of the statistics code book selector 33 is a table which matches the element (code vector) of the integrated code book 321 with the code book 34 for reference based on the sound classification flag 16.

[0095] Thus, there is the curtailment effect of storage capacity rather than it provides two or more code books individually by adoption of an integrated code book.

[0096] Moreover, although one code book for reference (subset) was limited from the integrated code book in the example of drawing 19 , it is also possible to limit two or more code books for reference. This example is shown in drawing 20 . In the statistics code book selector 33, two or more code books 34a, ..., 34k for reference are outputted. This corresponds, when using two or more candidate code vectors in the 3rd or the 6th example.

[0097] In the above explanation, he was not conscious about especially the overlap between the elements (code vector) in two or more code books for reference. However, even when performing a sound classification, the boundary is ambiguous and cannot be separated completely. If there is no overlap in a subset supposing a sound classification mistakes by few differences, degradation of voice quality will be caused. On the other hand, if the subset of an integrated code book is made to overlap intentionally and is constituted as shown in drawing 21 , it is possible to make influence of the error of a sound classification into the minimum.

[0098] As mentioned above, even if it transposes two or more statistics code books of the 1st to 6th example to an integrated code book and carries out selection of the code book for reference to limitation of the subset of an integrated code book, it is clear that the same function is realizable. Furthermore, there is an effect which can cut down the storage capacity of a code book.

[0099]

[Effect of the Invention] According to this invention, the quality low bit

rate voice coding method of about 3.6 kbpses can be comparatively offered by the low throughput.

[Translation done.]